Forum Jeunes Chercheurs à Inforsid 2014*

Guillaume Cabanac [1], **Amira Derradji** [2], **Ali Jaffal** [3],
Jonathan Louëdec [1,4], **Gloria Elena Jaramillo Rojas** [5]

1. *Institut de Recherche en Informatique de Toulouse – UMR 5505*
   Université de Toulouse
   guillaume.cabanac@univ-tlse3.fr, jonathan.louedec@irit.fr
2. *Laboratoire d'Informatique de Grenoble – UMR 5217*
   Université Grenoble Alpes & ARCAN SYSTEMS, Bron
   amira.derradji@imag.fr
3. *Centre de Recherche en Informatique – EA 1445*
   Université Paris 1 Panthéon-Sorbonne
   ali.jaffal@malix.univ-paris1.fr
4. *Institut de Mathématiques de Toulouse – UMR 5219*
   Université de Toulouse
   jonathan.louedec@irit.fr
5. *Laboratoire d'Informatique de l'Université de Pau et des Pays de l'Adour – EA 3000*
   Mont-de-Marsan
   gloriaelena.jaramillorojas@univ-pau.fr

RÉSUMÉ. *Le Forum Jeunes Chercheurs a été organisé lors du congrès Inforsid 2014 à Lyon. Il a accueilli 17 doctorants de première ou deuxième année travaillant dans le domaine des systèmes d'information. Ils ont rédigé un article et présenté leurs travaux lors d'une session plénière du congrès. Cet article coordonné par Guillaume Cabanac (en qualité d'organisateur du Forum) présente une sélection des quatre meilleures contributions au forum.*

ABSTRACT. *The Forum Jeunes Chercheurs was organised during the Inforsid 2014 congress held in Lyon, France. It hosted 17 first-year or second-year PhD students working in the Information Systems field. They wrote a paper and gave a talk about their research during a plenary session of the congress. This article coordinated by Guillaume Cabanac (as the organiser of the Forum) features a selection of the four best contributions to the forum.*

MOTS-CLÉS : *Forum jeunes chercheurs, Inforsid, système d'information.*

KEYWORDS: *Forum jeunes chercheurs, Inforsid, information systems.*

DOI:10.3166/ISI.20.2.119-143

*. Les contributeurs de cet article sont listés alphabétiquement.

1. Introduction

La septième édition du Forum Jeunes Chercheurs a été organisée à Lyon le 22 mai 2014, dans le cadre du 32e congrès Inforsid. L'objectif de cette manifestation scientifique était double. Premièrement, il s'agissait de permettre aux jeunes chercheurs en première ou deuxième année de doctorat de présenter leur problématique de recherche et d'établir des contacts avec des équipes travaillant sur les domaines similaires ou connexes. Deuxièmement, le forum visait à offrir un aperçu des axes de recherche actuels (Collectif Inforsid, 2012) et ainsi élargir le champ des connaissances des jeunes chercheurs et des congressistes au sens large.

Les actes du Forum Jeunes Chercheurs (Cabanac, 2014) recueillent les 17 articles soumis par des doctorants issus de divers laboratoires de recherche en France, au Maroc et en Tunisie (figure 1). Ces articles ont fait l'objet d'une présentation lors de la session plénière du congrès consacrée au forum. Durant toute la durée du congrès, l'exposition des posters préparés par les doctorants a également offert des opportunités de rencontres et de discussions avec les chercheurs de la communauté des systèmes d'information.



Figure 1. Répartition géographique des affiliations des 17 doctorants ayant participé au Forum Jeunes Chercheurs à Inforsid 2014

Le nombre et la qualité des articles ainsi que la variété des thématiques abordées par les doctorants participant au Forum Jeunes Chercheurs sont autant d'éléments qui attestent de la vigueur des recherches dans le domaine des systèmes d'information.

Cet article met à l'honneur les travaux de quatre jeunes chercheurs ayant participé au forum. Ces contributions ont été sélectionnées sur la base de la qualité de l'exposé et de l'article de quatre pages publié dans les actes du Forum Jeunes Chercheurs. Les doctorants sélectionnés ont été invités à détailler leur travail de doctorat sur six pages, ces dernières étant assemblées dans les sections suivantes. En section 2, Amira Derradji détaille la construction de protocoles de soins auto-adaptatifs pour le suivi des maladies. En section 3, Ali Jaffal traite de l'analyse formelle de concepts pour la gestion de contexte en informatique ubiquitaire. En section 4, Jonathan Louëdec présente l'application d'algorithmes de bandits pour les systèmes de recommandation. En section 5, Gloria Elena Jaramillo Rojas considère des accords de niveau de service et la modélisation de la réputation pour le contrôle d'usage. Le présent article a été coordonné par Guillaume Cabanac, organisateur du Forum Jeunes Chercheurs 2014.

2. Construction de protocoles de soins auto-adaptatifs pour le suivi des maladies : contribution d'Amira Derradji

Amira Derradji est doctorante de première année au LIG de Grenoble [1], sous la direction d'Agnès Front et de Christine Verdier.

2.1. Contexte

Le transfert de soins de l'hôpital vers le domicile du patient a vu le jour depuis de nombreuses années avec la naissance du concept américain « HomeCare ». En France en particulier, cette solution a été adoptée pour diverses raisons démographiques (le vieillissement de la population), économiques (le coût élevé d'hospitalisation), sociales (l'engorgement des hôpitaux, la volonté des patients de se soigner chez eux), médicales (l'apparition de nouveaux traitements) et politiques (priorité importante du gouvernement français étayée par différents décrets, circulaires, etc.). En revanche, il reste à l'heure actuelle un enjeu majeur pour assurer la continuité des soins et améliorer la qualité de vie du patient à l'extérieur du milieu hospitalier.

Notre recherche s'inscrit dans le cadre de la prise en charge médicale à domicile de personnes atteintes de maladies chroniques. Ces maladies se caractérisent par une longue phase pathologique et un profil évolutif et peuvent engendrer une perte d'autonomie chez les patients atteints, provoquant ainsi une diminution de leurs capacités. Des solutions technologiques peuvent permettre d'endiguer la perte d'autonomie. Depuis plusieurs années, diverses pistes de prise en charge ont vu le jour pour mieux accompagner le patient au quotidien. L'Hospitalisation À Domicile (HÀD), le Service de Soins

1. http://www.liglab.fr

Infirmiers À Domicile (SSIÀD) ou encore le Maintien À Domicile (MÀD) permettent au patient d'être soigné dans son environnement familier et aux professionnels de santé d'effectuer un suivi régulier à distance *via*, parfois, des dispositifs installés à domicile mais le plus souvent, des cahiers de liaison.

À domicile, chaque patient est amené à suivre un protocole de soins défini et personnalisé par son équipe soignante. D'un point de vue informatique, nous considérons ce protocole comme un processus. Un processus est un ensemble d'activités structuré et ordonné chronologiquement, réalisé à l'aide de ressources matérielles et immatérielles dans le but de réaliser un objectif. Par exemple, le protocole de soins de la surveillance de dialyse peut être considéré comme une succession d'activités réalisées pendant la séance de dialyse par le patient et éventuellement ses aidants à l'aide d'un dispositif médical (le dialyseur) et dont le résultat final est le déroulement correct de la dialyse.

Cependant, ce protocole doit évoluer avec (i) l'évolution de la maladie chronique dans le temps et (ii) les actions réalisées ou subies par le patient pendant l'application du protocole à domicile. La définition du protocole n'est donc pas définitive, son exécution se déroule dans un environnement dynamique, ce qui nécessite de l'adapter en fonction des changements afin d'assurer un suivi personnalisé des soins au patient.

De plus, nous prenons en compte dans cette recherche l'expertise du patient. Un patient expert peut être défini comme une personne qui connaît parfaitement sa maladie chronique (le vécu au quotidien, les traitements médicaux, etc.), a acquis et développé des connaissances fortes et a pu tirer une expertise très utile pour lui-même et l'équipe médicale. Il devient alors une partie intégrante de ses propres soins. Nos travaux visent donc à concevoir des protocoles de soins auto-adaptatifs intégrant l'expertise du patient.

2.2. État de l'art

De nombreuses solutions permettant la prise en charge des patients à domicile sont apparues au cours des dernières décennies. Certaines sont centrées sur *la gestion des activités de soins réalisées par les professionnels de santé*, notamment les travaux de (Bricon-Souf *et al*., 2005), menés au CERIM où les auteurs ont proposé une architecture intégrant les connaissances cognitives de la prise en charge. La solution *Athome* de notre collaborateur industriel *Arcan Systems* [2] s'inscrit également dans la coordination de l'ensemble des acteurs à domicile. D'autres solutions sont orientées *processus de soins*, elles sont basées sur des services de télé-santé (télé-consultation : service GLUCONET (Perréal, 2003), télé-assistance : projet TelePat (Lacombe *et al*., 2006), télé-surveillance : projet e-vital (Sakka *et al*., 2004) et télé-expertise : réseau TELURGE (Desrues, 2008)). Deux projets sont interactifs avec le patient : les projets de télé-consultation et de télé-surveillance à domicile. Les données recueillies sont issues de dispositifs médicaux (implantés ou non) et se présentent sous forme de mesures (glycémie, pouls, etc.),

2. Arcan Systems : http://www.arcan.fr

de signaux physiologiques (électrocardiogramme, électro-encéphalogramme, etc.) ou d'imagerie fixe et dynamique (image, vidéo) (Collectif DGOS, 2012).

S'intéresser aux *workflows* adaptatifs suppose d'étudier l'adaptation dans les processus. Plusieurs approches d'adaptation des modèles de processus ont été proposées dans la littérature (Weber *et al.*, 2008 ; Weske, 2012 ; Andonoff *et al.*, 2012). Selon la classification proposée par Andonoff *et al.* (2012), certaines approches sont basées sur l'adaptation pendant la phase de modélisation du processus (par exemple l'approche par conception et l'approche par intention et contexte) et d'autres, pendant la phase d'exécution (par exemple l'approche par intégration de protocole, l'approche par intention et contexte, l'approche par cas, l'approche par contraintes et l'approche par évolution). Dans (Andonoff *et al.*, 2012), les approches sont classées selon trois formes principales d'adaptation : l'adaptation par flexibilité, qui a pour objectif de retarder la modélisation ou l'exécution du processus, l'adaptation par déviation, qui traite les imprévus pendant la phase d'exécution et l'adaptation par évolution, qui analyse l'exécution afin d'apporter des améliorations a posteriori. Ces trois formes d'adaptation sont inspirées de plusieurs propositions de typologies existantes (Nurcan, 2008 ; Schonenberg *et al.*, 2008 ; Weber *et al.*, 2009).

2.3. Problématique

L'idée générale de notre travail consiste à mettre en évidence l'*expertise et le savoir du patient* afin d'améliorer son suivi médical loin de l'hôpital. Notre objectif est de fournir au patient un outil informatique lui permettant de signaler tout événement spécifique prévu et imprévu (cf. figure 2). L'équipe médicale pourra ensuite interpréter la situation, et mettre en œuvre une action ciblée. Cet outil présentera également un intérêt réel pour l'amélioration de la connaissance médicale car toutes les informations renseignées par le patient seront recueillies et restituées pour répondre à de nouvelles situations pathologiques et aider le professionnel de santé à mieux s'informer.

Plusieurs défis devront être résolus tels que : comment filtrer les événements imprévus pour ne retenir que ceux qui ont un sens pour le protocole ? comment automatiser le protocole de soins et l'adapter en fonction des événements perturbateurs ? comment permettre l'auto-enrichissement du protocole de soins ? Notre réponse sera basée sur les techniques d'adaptation des processus à l'exécution.

2.4. Étude et modélisation d'un processus de soins

Jusqu'à présent, nous avons dirigé nos recherches autour de deux axes principaux. L'un concerne les processus traditionnels (statiques), définis dans un environnement plus ou moins stable. L'autre concerne les processus adaptatifs (dynamiques), qui ne peuvent pas être complètement spécifiés à l'avance à cause des changements imprévus qu'ils peuvent subir aussi bien dans la phase de modélisation qu'à l'exécution.

Dans le premier axe, nous nous sommes intéressés à la modélisation d'un processus statique correspondant au processus de soins « surveiller la dialyse » selon différentes

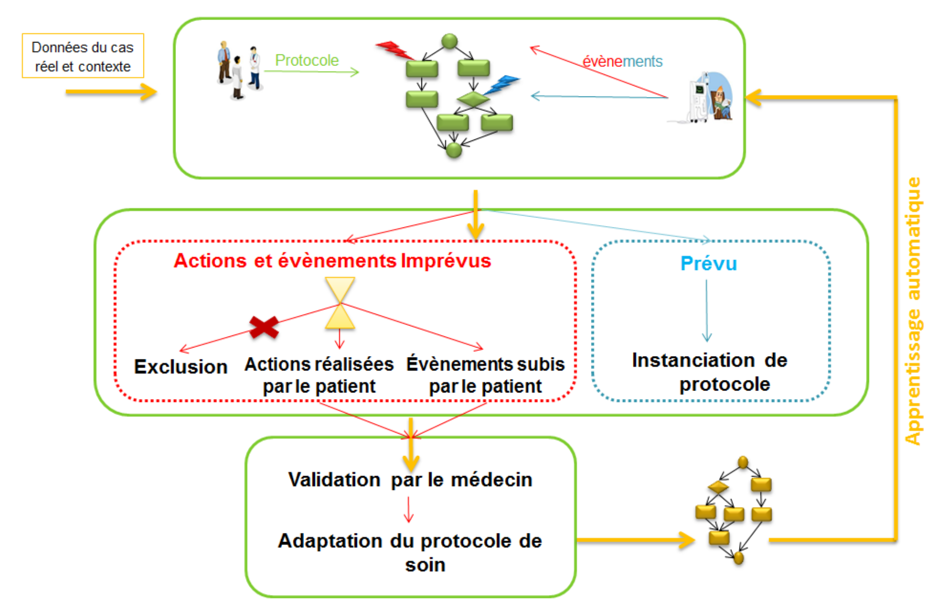


Figure 2. Filtrage des événements

approches de modélisation à savoir UML (Morley *et al.*, 2006), BPMN (OMG, 2010), le diagramme EPC (Davis, 2001), le logigramme, les schémas de processus (Brandenburg, Wojtyna, 2006) ainsi que les diagrammes de description (Noyé, 1997). Deux approches nous ont permis de représenter les divers événements correspondant aux besoins du patient : BPMN et EPC. À cela s'ajoute la possibilité de modéliser les sous-processus de soins sous BPMN. Le langage BPMN semble le plus approprié pour la modélisation des événements prévus liés aux processus statiques (cf. figure 3) mais il est mal adapté à des utilisateurs finaux (patient, médecin).

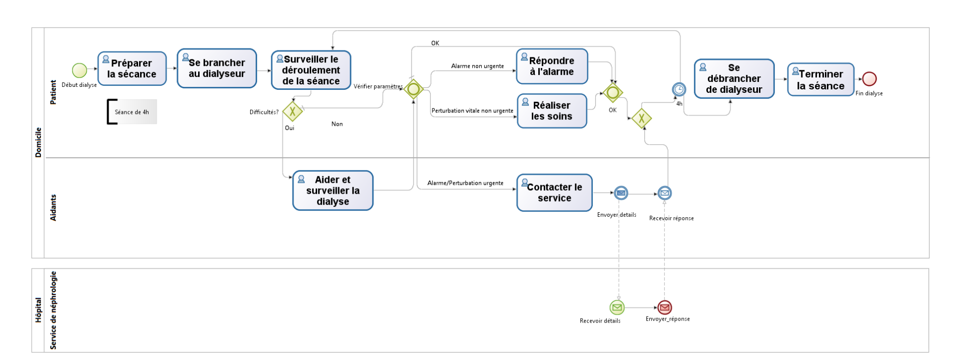


Figure 3. Protocole de soins « surveiller la dialyse » sous BPMN

Dans le deuxième axe, un approfondissement a été mené sur les processus dynamiques où nous avons abordé les trois principaux types d'adaptation et les différentes

approches d'adaptation des processus. Un processus dynamique peut être modifié par des événements prévus ou imprévus comme nous le détaillons dans la suite.

2.5. Événements et actions prévus et imprévus

L'adaptation des processus de soins nécessite (i) le filtrage intelligent des événements et des actions. Un événement est défini comme une occurrence temporelle pouvant changer le protocole de soins. Suite à un événement, une action peut être réalisée par le patient. Par exemple, pour l'événement « avoir des démangeaisons de la peau », l'action prévue consiste à « appliquer un baume apaisant ». Pour chaque événement imprévu, une action imprévue peut être réalisée. Par exemple pour l'événement imprévu « problèmes cardiaques », le patient réalisera l'action imprévue « appeler le médecin ». Une fois le filtrage réalisé, une adaptation du protocole en fonction des événements imprévus pertinents est effectuée (après accord de l'équipe médicale). Enfin, les adaptations réalisées seront intégrées dans le système afin d'être en conformité avec le vécu du patient si ces mêmes besoins surviennent de nouveau.

2.6. Perspectives

Dans la suite de notre travail, certaines solutions candidates aux défis identifiés seront étudiées telles que les *ontologies* qui permettent de représenter les connaissances du domaine afin de les manipuler et leur associer un raisonnement. Une approche d'adaptation de processus de soins sera appliquée d'une manière (semi)-automatique en utilisant les techniques d'intelligence artificielle comme le *raisonnement par cas*. Enfin, face à la question de l'*auto-adaptabilité*, nous ferons appel à des techniques issues de l'intelligence artificielle comme par exemple l'*apprentissage automatique*.

3. Analyse formelle de concepts pour la gestion de contexte en informatique ubiquitaire : contribution d'Ali Jaffal

Ali Jaffal est doctorant de première année au CRI de Paris [3], sous la direction de Bénédicte Le Grand et de Manuele Kirsch Pinheiro.

3.1. Contexte

Ce sujet de recherche s'inscrit dans le domaine des systèmes d'information pervasifs. Ceux-ci étendent les systèmes d'information traditionnels en s'adaptant aux conditions dans lesquelles ils sont utilisés. L'objectif est de gérer et d'analyser les informations liées au contexte d'un utilisateur, disponibles grâce à l'informatique ubiquitaire. D'énormes volumes de données très diverses sont en effet collectés par des capteurs, fournissant par exemple des indicateurs sur la localisation géographique d'un individu,

3. http://www.univ-paris1.fr/centres-de-recherche/cri

les conditions météorologiques, etc. Les collectes de données de ce type se multiplient, par exemple dans le cadre des villes intelligentes. Ce contexte « environnemental » peut être enrichi d'autres informations relatives aux utilisateurs, par exemple, leurs activités ou leur réseau social. Ceux-ci peuvent aussi être eux-mêmes fournisseurs d'indicateurs lorsqu'ils renseignent leur profil. L'objectif des travaux est de pouvoir anticiper les actions futures d'un utilisateur en fonction des évolutions de ce contexte ou de lui faire des recommandations d'actions pertinentes.

3.2. État de l'art

Apparue au début des années quatre-vingt-dix à travers notamment les travaux de Schilit et Theimer (1994), la sensibilité au contexte a été définie comme la capacité d'une application à découvrir et à réagir aux modifications de l'environnement de l'utilisateur (Dey, 2001). Au centre de ces systèmes se trouvent donc les mécanismes de gestion de contexte (Dey, 2001 ; Conan *et al.*, 2007 ; Brézillon, 2003), responsables de l'acquisition et de la représentation de l'information de contexte à partir de l'environnement. C'est à partir de cette information que les processus d'adaptation caractéristiques des systèmes sensibles au contexte peuvent prendre place.

La notion de contexte peut être vue comme « l'ensemble des caractéristiques de l'environnement physique ou virtuel qui affecte le comportement d'une application et dont la représentation et l'acquisition sont essentielles à l'adaptation des informations et des services » (Gensel *et al.*, 2008). Il peut s'agir, par exemple, de l'observation de la localisation géographique d'un utilisateur (typique des systèmes de géolocalisation), ou encore de l'observation du terminal utilisé (capacité de la mémoire ou état de la batterie) pour l'adaptation du contenu offert (par ex. : les points d'intérêts les plus proches) ou même du fonctionnement de l'application (par ex. : réduction de la luminosité pour économiser la batterie). L'information de contexte est complexe, du fait de son caractère dynamique et incomplet, voire incertain (Chalmers *et al.*, 2004 ; Vanrompay *et al.*, 2011). Par ailleurs, l'acquisition des informations dans un tel environnement est sujette à des imprécisions, du fait notamment de limitations des équipements (capteur défaillant) ou des techniques de collecte utilisées.

La notion de recommandation associée à la notion de contexte n'est pas très détaillée dans la littérature et est souvent ambiguë. Pignotti *et al.* (2004) parlent de recommandation adaptative et présentent le contexte comme un élément majeur pour ce type d'adaptation. Par contexte, les auteurs entendent la date, l'emplacement et l'historique des services utilisés par l'utilisateur. Wen et He (2006) introduisent un système de recommandation personnalisée de services, qui comporte un ensemble d'outils permettant de trouver des informations ou des services intéressants pour l'utilisateur. Chaari *et al.* (2009) relient la recommandation à la personnalisation et à l'adaptation. Le concept de base de ce travail est ainsi de générer automatiquement du code pour l'interface utilisateur afin de fournir des services personnalisés en adaptant les offres, en recommandant certains produits, ou en offrant un accès facile à quelques services en fonction de la localisation de l'utilisateur.

Un autre exemple est donné par Najar *et al.* (2012), qui proposent un mécanisme de prédiction cherchant à anticiper les services à invoquer pour un utilisateur en fonction de son contexte. Sigg *et al.* (2010) proposent, quant à eux, de compléter la description d'une situation avec les éléments de contexte manquants à partir de l'analyse des contextes précédemment observés. Ces travaux utilisent majoritairement des méthodes statistiques et des mesures de similarité pour effectuer une classification des données.

3.3. Problématique

Bien que les approches mentionnées fournissent des résultats intéressants, elles présentent certaines limites. Les méthodes de classification sous-jacentes ne permettent notamment pas de générer des classes recouvrantes. En d'autres termes, les éléments de contexte ne peuvent pas appartenir simultanément à plusieurs classes, alors que, dans la pratique, un même élément de contexte peut caractériser différentes situations. Par ailleurs, pour être efficaces, ces méthodes requièrent un volume important de données, qui ne sont pas toujours disponibles. Enfin, la justification des classes identifiées n'est pas forcément compréhensible pour les utilisateurs finaux, ce qui peut conduire à des erreurs d'interprétation et donc de mauvaises recommandations.

Les questions auxquelles nous souhaitons répondre sont notamment :

– Comment identifier les informations de contexte pertinentes ?

– Dans quel contexte les actions sont-elles réalisées ?

– Quel est l'impact de ces informations sur le choix (et donc la recommandation) d'une action par (ou pour) un utilisateur ?

3.4. Actions réalisées

L'objectif de cette thèse est d'exploiter au mieux les données disponibles afin de comprendre le comportement de l'utilisateur et éventuellement faire des recommandations. Nous proposons une méthodologie qui repose sur l'Analyse Formelle de Concepts (AFC) (Priss, 2006 ; Wille, 2005) et les règles d'association (Agrawal *et al.*, 1993 ; Agrawal, Srikant, 1994) pour la gestion de contexte à des fins de personnalisation et de recommandation. Cette méthode permet de découvrir des relations non triviales entre les éléments de contexte et des actions apparemment décorrélées et d'évaluer l'impact des éléments de contexte sur les actions de l'utilisateur. L'AFC est une méthode mathématique pour l'analyse de données qui permet le regroupement conceptuel par des treillis de concepts, également appelés treillis de Galois. Grâce à l'AFC, nous pouvons organiser les informations contextuelles observées sur des classes pouvant se recouvrir, quelle que soit la granularité de l'information. Contrairement à d'autres méthodes d'analyse, l'AFC permet de retrouver une structuration naturelle des données, en associant les actions aux éléments de contexte observés. Les règles d'association permettent, quant à elles, d'identifier les relations de causalité.

Notre méthodologie, illustrée sur la figure 4, se concentre sur la relation entre l'environnement et le système de gestion de contexte. D'une part, l'environnement est défini par les différents éléments de contexte observés : le contexte temporel, la localisation géographique, l'interaction utilisateur-application ou encore l'activité courante de l'utilisateur. D'autre part, la gestion de contexte est effectuée en trois étapes : 1) collecte et mise en forme des éléments de contexte provenant de l'environnement de l'utilisateur ; 2) AFC et construction de treillis de Galois pour structurer les éléments de contexte ; 3) extraction de règles d'association pour évaluer l'impact des éléments de contexte sur les actions de l'utilisateur.

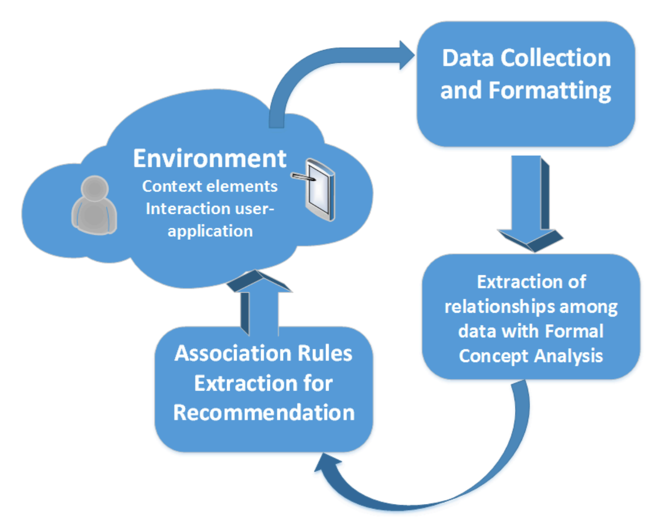


Figure 4. Méthodologie pour la gestion et l'analyse de contexte

La méthodologie de gestion conceptuelle de contexte proposée ici a été appliquée dans une étude de cas réalisée à partir des observations de contexte d'un utilisateur réel dans son utilisation quotidienne d'une tablette tactile.

Le point de départ de l'analyse est la collecte et la sauvegarde des données liées aux actions et à l'environnement de l'utilisateur (applications utilisées, diverses localisations géographiques, différents réseaux de connexion, etc.).

La construction d'un treillis de Galois s'effectue à partir d'un contexte formel $K(G, M, I)$ qui regroupe les *objets* (applications) d'un ensemble G et les *attributs* (éléments de contexte) d'un ensemble M qui leur sont associés. Par exemple, l'activité A2, dans la figure 5, a été effectuée dans des contextes C2 et C3. Un concept formel est une paire (O, A) où O est un ensemble d'objets et A l'ensemble des attributs communs à ces objets. L'ensemble de tous les concepts muni de l'ordre $\leqslant$ forme le treillis de concepts. Par exemple, l'un des concepts du treillis de la figure 5 montre que les activités A1 et A3 ont en commun les éléments de contexte C1 et C3 (les étiquettes des attributs sont héritées par les concepts inférieurs dans le treillis, alors que les étiquettes des objets sont héritées par les concepts supérieurs).

Attributs / Objets	C1	C2	C3	C4
A1	1	0	1	0
A2	0	1	0	1
A3	1	0	1	0
A4	1	0	0	1

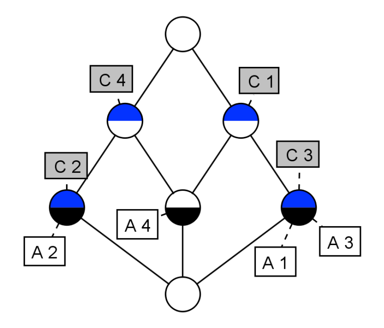


Figure 5. Exemple de contexte formel et son treillis

L'extraction de règles d'association a pour but de découvrir des relations significatives entre attributs binaires extraits des bases de données. Par rapport à d'autres techniques classiquement utilisées pour la recommandation, notamment les techniques basées sur le contenu, les règles d'association ne requièrent pas le calcul d'une mesure de similarité. Ceci est particulièrement intéressant dans notre contexte où les éléments de contexte et les actions ne sont pas forcément comparables.

Une règle d'association est définie comme une implication entre deux ensembles. Elle est de la forme « $R : X \Rightarrow Y$ tel que $X \subseteq I$, $Y \subseteq I$ et $X \cap Y = \emptyset$ ». La règle R est dite basée sur l'ensemble fréquent $X \cup Y$ et les deux ensembles X et Y sont appelés, respectivement, *prémisse* et *conclusion* de R. Afin de vérifier la validité d'une règle d'association R, deux mesures sont communément utilisées, le support et la confiance. Le tableau 1 suivant représente un échantillon de prédictions pour notre cas d'étude (avec divers indices de confiance). $R1$ est une prédiction de confiance égale à 100 %, et signifie que si l'utilisateur emploie l'application *Maps* lorsqu'il est connecté au réseau *Network1* pendant la nuit, il emploie par la suite l'application *9GAG*.

Tableau 1. Exemples de prédictions

	Prédictions	Conf
R1	[t_Night, Network1, Maps] ⇒ [9GAG]	100,00 %
R2	[E-mail, Gmail, t_Evening , Network1] ⇒ [Google+]	83,33 %
R3	[t_Afternoon, Location1] ⇒ [Chrome]	61,11 %
R4	[t_Night, Location1, Network1] ⇒ [LaStampa.it]	45,00 %
R5	[t_Evening, Location1] ⇒ [Gmail]	40,90 %
R6	[t_Night, Location1, Chrome] ⇒ [Facebook]	35,71 %

3.5. Perspectives

L'approche et la solution proposées permettent d'envisager de nombreuses perspectives. La première consiste à concevoir des mécanismes pour automatiser les interprétations et les recommandations associées. Nous exploiterons pour cela les concepts

générés par le treillis de Galois et tous les liens entre ces concepts, afin d'en déduire de manière automatique une interprétation et des recommandations.

Nous avons jusqu'ici utilisé un nombre limité d'éléments de contexte (localisation géographique, créneau horaire, réseau d'accès et mémoire du terminal). Nous étudierons dans l'avenir la pertinence d'autres types d'éléments de contexte afin d'étendre notre approche.

Nous souhaitons également appliquer notre méthodologie à d'autres cas d'étude, et réaliser une évaluation des recommandations effectuées auprès des utilisateurs. À plus long terme, nous souhaitons effectuer une gestion de contexte en temps réel, qui posera des problèmes de passage à l'échelle du fait du volume de données à analyser et aux contraintes temporelles. Une piste possible consiste à utiliser des approches distribuées pour la construction et la mise à jour des treillis de Galois.

4. Application d'algorithmes de bandits pour les systèmes de recommandation : contribution de Jonathan Louëdec

Jonathan Louëdec est doctorant de première année à l'IRIT et à l'IMT de Toulouse [4], sous la direction de Max Chevalier, Aurélien Garivier et Josiane Mothe.

4.1. Contexte

Les systèmes de recommandation (SR) ont pour objectif de proposer à l'utilisateur des items (documents, objets, films, musiques, informations, etc.) susceptibles de l'intéresser. Deux approches principales sont mises en œuvre dans les SR : le filtrage basé sur le contenu (FBC) recommande à un utilisateur des items similaires à ceux qu'il a déjà aimés par le passé et le filtrage collaboratif (FC) recommande les items appréciés par les utilisateurs qui ont auparavant fait des choix similaires à ceux de l'utilisateur. D'autres approches existent comme, par exemple, le filtrage démographique qui se base sur ce que l'on sait de l'utilisateur (âge, données démographiques, sexe, etc.) ; le filtrage communautaire qui utilise les décisions faites par les contacts de cet utilisateur (cette méthode est notamment utilisée dans les SR sociaux).

4.2. État de l'art

Un des enjeux des SR est d'intégrer de nouveaux items et de nouveaux utilisateurs qui apparaissent au cours du temps. Les deux types d'approche, FC et FBC souffrent de limites bien identifiées lors de cette intégration. Un premier problème est d'identifier quels items doivent être proposés à un nouvel utilisateur. Ce problème, connu sous le nom de « démarrage à froid utilisateur », est commun aux FC et FBC. Par ailleurs, il existe des problèmes spécifiques à chaque type d'approche. Le départ à froid côté item

4. http://www.irit.fr et http://www.math.univ-toulouse.fr

(à qui proposer un nouvel item ?) ne concerne que le FC puisque le FBC proposera un nouvel item aux utilisateurs qui ont aimé un item similaire. Cependant, si le contenu d'un item est peu décrit, le FBC ne pourra pas établir de similarité, tandis que le FC pourra proposer ce nouvel item en fonction de l'information obtenue par les actions passées des utilisateurs (Montaner, 2003). Ainsi, la combinaison entre FC et FBC peut permettre de corriger les limites spécifiques de chaque approche : actuellement, les SR les plus efficaces sont effectivement basés sur une approche hybride (Ricci *et al.*, 2011).

Koren (2010) souligne qu'un SR doit par ailleurs être capable de prendre en compte l'évolution des besoins de l'utilisateur : c'est la définition d'un SR dynamique. L'utilisation d'une fenêtre temporelle, permettant de prendre en compte uniquement les événements les plus récents et la mise en place de méthodes de détection de rupture sont utilisées pour aborder ce problème. Une autre approche est possible : au lieu de considérer uniquement les informations passées les plus récentes, les informations passées peuvent être pondérées : plus une information est ancienne, moins elle aura d'importance dans le calcul des estimateurs.

4.3. Problématique

Quelle que soit l'approche de recommandation retenue, à un instant t de l'interaction avec le système, se pose la question de la sélection des items à proposer à l'utilisateur à l'instant $t+1$. Même si les interactions des différents utilisateurs avec le SR jusqu'à l'instant t sont connues, de nouveaux événements sont susceptibles d'apparaître (nouvel utilisateur, nouveau document, évolution des goûts de l'utilisateur), une méthode statique de type FBC uniquement est donc mal adaptée ; la stratégie de recommandation doit pouvoir s'adapter à la non-stationnarité des flux de données.

Une stratégie doit donc être mise en œuvre pour acquérir de l'information sur les utilisateurs ainsi que sur l'ensemble des items (exploration) tout en assurant que le SR donnera de bons résultats (exploitation). Cette problématique est connue sous le nom du dilemme exploration/exploitation. Une stratégie efficace doit également prendre en compte le passage à l'échelle et rester performante même en présence de gros volumes de données (items, utilisateurs, interactions). Cela pose deux problèmes : comment stocker l'information? Comment optimiser le temps d'acquisition ou d'apprentissage?

Des modèles mathématiques permettent de rendre compte de ce dilemme entre exploration et exploitation : les « modèles de bandits ». On se propose de s'inspirer des travaux sur ces modèles pour construire des stratégies efficaces et dynamiques. Cette approche met en œuvre le dilemme exploration/exploitation en offrant des capacités d'apprentissage qui sont utilisées depuis des décennies dans plusieurs autres domaines : essais cliniques, « Yield management » (gestion des prix des tickets de train), gestion des tâches, routage de paquets internet, intelligence artificielle, théorie des jeux. . .

4.4. Actions réalisées

Fin 2013, le moteur de recherche russe Yandex a proposé un challenge dont l'objectif était de réordonner les dix premiers résultats de 797 867 requêtes soumises par des

utilisateurs différents [5]. Nous disposions de plus de 34 millions de requêtes passées et des actions utilisateurs liées : clics, temps passé sur la page cliquée. Chaque terme et chaque site web était remplacé par un identifiant numérique, il n'était donc pas possible de prendre en compte la sémantique des différents éléments. Les actions passées des utilisateurs constituaient les seuls éléments dont nous disposions pour résoudre le problème. Dans un premier temps nous avons dû mettre en place une structure permettant de stocker et de traiter rapidement ces données. Nous avons fait le choix d'utiliser MongoDB, un système de gestion de base de données NoSQL orientée documents. Ce type de système est reconnu pour être efficace avec de grands volumes de données. *Via* un script Python, l'insertion en base des données a duré deux heures. Une fois les index créés, la base occupait 117 Go.

L'ordonnancement initial fourni par les organisateurs obtenait déjà un score [6] élevé ; il est apparu difficile de dépasser ce score sans réutiliser le classement d'origine. Nous avons donc proposé plusieurs stratégies de réordonnancement. Celle présentée en Algorithme 1 a permis d'obtenir les meilleures performances. Cette configuration nous a permis d'apprendre le nouvel ordonnancement en deux heures environ. En utilisant cet algorithme, nous avons obtenu la 33e place sur 196 participants.

Algorithme 1 : Méthode mise en œuvre pour le ré-ordonnancement

pour *chaque requête R pour laquelle les résultats doivent être réordonnés* **faire**
  pour *chaque document D de la liste réponse de R* **faire**
    Récupérer le nombre de clics de l'utilisateur sur D avec au moins un terme de R dans la requête dans les sessions passées (apprentissage)
  fin
  Réordonner les documents en fonction de leur score, conserver l'ordre de départ si ex-æquo
fin

L'utilisation du contexte utilisateur a permis d'améliorer l'ordonnancement initial, confortant ainsi notre intérêt pour sa prise en compte dans notre problématique.

4.5. Actions futures

Lors de cette première expérimentation, l'ensemble des données était disponible *a priori*, ce qui nous plaçait dans un cadre d'apprentissage supervisé non dynamique. Ce cas de figure n'est pas complètement réaliste pour un SR. Ainsi, la suite de nos travaux s'effectuera dans un autre cadre d'apprentissage : les données arrivent progressivement et la décision doit être prise au même instant. Ce cadre est appelé « apprentissage par renforcement » (Sutton, Barto, 1998). À chaque interaction avec l'utilisateur, le système récupère les résultats, nous les appellerons récompenses. L'objectif d'un tel algorithme est de minimiser la différence entre la somme des récompenses obtenues si la solution

5. http://www.kaggle.com/c/yandex-personalized-web-search-challenge
6. NDCG : https://www.kaggle.com/wiki/NormalizedDiscountedCumulativeGain

optimale est utilisée, et la somme des récompenses obtenues *via* l'algorithme évalué. Cette quantité est appelée le regret.

Ainsi le cadre d'un apprentissage par renforcement est une piste prometteuse pour les SR. Les utilisateurs se connectent « séquentiellement » au système, de nombreux items pouvant potentiellement être recommandés sont disponibles. À chaque instant t, un certain nombre d'items sont proposés à l'utilisateur qui interagit avec le système. L'objectif du SR n'est pas de connaître la pertinence de chaque item parfaitement, mais plutôt que les recommandations fournies conduisent à un maximum de clics. Un item qui ne conduit pas à un clic (qui est donc considéré comme non pertinent) devra être mis de côté rapidement (exploitation). Cependant, il est important qu'il ait été tout de même proposé suffisamment de fois de sorte qu'il soit écarté avec un fort niveau de confiance (exploration). C'est ici que le regret défini précédemment prend tout son sens. En effet le regret va augmenter à court-terme car nous voudrons tester suffisamment de fois chaque item, mais il diminuera à plus long terme lorsque les items non pertinents seront écartés pour conserver uniquement les plus pertinents. Cette mesure est particulièrement adaptée aux méthodes prenant en compte le dilemme exploration/exploitation.

Les algorithmes de bandits sont connus pour proposer des solutions à ce dilemme, voir (Bubeck, Cesa-Bianchi, 2012). Cette approche est séquentielle: à partir des t premiers résultats, une décision est prise à l'instant $t + 1$. Dans cette approche chaque item possède un estimateur. À chaque instant t un ou plusieurs estimateurs sont mis à jour. Certains algorithmes de bandits, UCB *(Upper Confidence Bound)* par exemple (Bubeck, Cesa-Bianchi, 2012), utilisent des bornes supérieures de confiance comme estimateurs. La borne supérieure de confiance d'UCB de l'item j est calculée *via* l'équation présentée dans Algorithme 2.

Algorithme 2 : UCB1 (Upper Confidence Bound)

Initialisation : Proposer une fois les n items
pour *chaque instant* $t > n$ **faire**
  Proposer l'item I_t qui maximise $\bar{x_j} + \sqrt{\frac{2 \ln t}{n_j}}$ avec :
  $\bar{x_j}$ la récompense moyenne obtenue pour l'item j
  n_j le nombre de fois où l'item j a été proposé
  Récupérer l'information induite par l'action utilisateur
  Mettre à jour la borne supérieure de confiance associée à I_t en fonction du résultat obtenu
fin

Plus un item est proposé, plus l'estimateur converge vers une valeur proche de la valeur réelle du paramètre. Si un nouvel item apparaît, la borne supérieure de confiance sera très grande. Une conséquence de cette approche optimiste est qu'un nouvel item sera proposé un certain nombre de fois dès son apparition pour estimer sa valeur. Cette stratégie peut être une approche intéressante pour aborder le problème de départ à froid coté item. Le FBC peut également aborder ce problème mais en contrepartie

d'autres aspects liés aux flux de données sont plus difficilement abordables, comme le départ à froid côté utilisateur. Concernant la prise en compte du vieillissement d'un item en utilisant UCB, il faudrait adapter la méthode de calcul de la borne supérieure pour détecter cette évolution. Pour cela l'utilisation d'une fenêtre temporelle et la pondération de l'importance des items en fonction du temps peut être étudiée. L'inconvénient de ces approches est le fait que les informations sur toute la fenêtre temporelle doivent être stockées dans le premier cas, les informations depuis l'origine du SR doivent être stockées dans le second cas. Une approche de type UCB n'utilisant pas ces méthodes peut stocker uniquement les estimateurs et le nombre de fois que chaque item a été proposé. La complexité de la méthode choisie pour un SR peut rendre plus difficile le passage à l'échelle.

Dans les versions les plus basiques, les algorithmes de bandits traitent chaque item de manière indépendante : un clic utilisateur sur un item modifiera uniquement l'estimateur de cet item. S'il y a de nombreux items à traiter, nous souhaitons inférer de l'information sur les items similaires à celui choisi afin de limiter le temps d'apprentissage. Lorsqu'un nouvel utilisateur apparaît, il s'agit d'acquérir un maximum d'informations en un minimum de temps pour estimer ses goûts. Actuellement les méthodes d'apprentissage actif (Rubens *et al.*, 2011) traitent ce problème en choisissant à chaque instant les items qui apporteront le plus d'informations. L'interaction des méthodes d'apprentissage actif avec celles des problèmes de bandits devrait permettre d'optimiser la partie exploration au profit de l'exploitation.

En recherche d'information, plusieurs chercheurs ont adapté les algorithmes de bandits. Li *et al.* (2010) proposent un algorithme linéaire de bandits, LinUCB, prenant en compte le contexte utilisateur, Dans le cadre du ré-ordonnancement d'items, Radlinski *et al.* (2008) proposent le « Ranked Bandits Algorithm ». Les auteurs mettent en place des algorithmes de bandits pour chaque position possible dans l'ordonnancement des résultats. Ils prouvent que sur le long terme, l'algorithme mis en place converge plus rapidement vers une solution acceptable qu'un algorithme basé uniquement sur la popularité des documents. La méthode d'évaluation proposée dans cet article reste tout de même très théorique. Kohli *et al.* (2013) se sont inspirés de cette approche et ont modifié notamment la façon dont le clic des utilisateurs est pris en compte, (« Independent Bandits Approach »). Pour comparer leur approche avec le « Ranked Bandits Algorithm », ils ont mis en place un cadre applicatif plus réaliste. Ces expérimentations se basent sur une partie du jeu de données MovieLens, contenant dix millions de notes réparties sur 10 681 films et 71 567 utilisateurs. Dans ces deux approches, « Ranked Bandits » et « Independent Bandits », les rangs et les clics sont traités de manière indépendante.

Le cas idéal de l'indépendance n'est pas toujours réaliste. Prenons la recommandation de films pour exemple. Si un utilisateur clique sur « Star Wars I », il est possible d'induire de l'information sur un possible clic sur les autres films de cette hexalogie ou du même type. Dans ce cadre, les k documents les plus populaires peuvent être aimés par les mêmes utilisateurs, représentant la majorité. Mais si l'utilisateur ne fait pas partie de cette majorité, aucune des propositions ne le satisfera. Suivant ce constat,

nous ne voulons plus proposer les k documents les plus populaires, mais plutôt la combinaison de k documents qui provoquera un clic au moins chez un maximum d'utilisateurs, et pas seulement chez une majorité ayant des goûts similaires. Cette approche favorise un critère important dans les SR : celui de la diversité (Candillier *et al.*, 2012).

5. Accords de niveau de service et la modélisation de la réputation pour le contrôle d'usage : contribution de Gloria Elena Jaramillo Rojas

Gloria Elena Jaramillo Rojas est doctorante de deuxième année au LIUPPA sur le site de Mont-de-Marsan [7], sous la direction de Philippe Aniorté et Manuel Munier.

5.1. Contexte

La sous-traitance de services est devenue une nouvelle forme d'interaction pour les entreprises en lui permettant de se concentrer sur leur cœur de métier en profitant des avantages des architectures orientées services. La collaboration résultant de cette sous-traitance entraîne la création de services composés et de *workflows* complexes dans lesquels le client d'un service partage des informations sensibles avec d'autres organisations externes et indépendantes. Dans un tel contexte inter-organisationnel, plusieurs défis restent à être traités par la communauté :

1. le partage explicite de données entre les différents partenaires engendre une perte de contrôle sur les données passées au(x) fournisseur(s) de service(s) ;

2. la complexité de la chaîne de processus nécessite de définir des stratégies de gestion de la confiance et de gestion des responsabilités (et des pénalités) en cas de mauvaise exécution des tâches prévues ;

3. la création de *workflows* dynamiques permet de poursuivre ces interactions inter-organisationnelles malgré des comportements défectueux tout en préservant le niveau de service requis par le client ;

4. les partenaires du *workflow* doivent prouver dans leurs comportements qu'ils respectent les politiques établies, mais agissant comme des boîtes noires ils doivent garder leur autonomie.

Pour faire confiance au fournisseur de services, l'accord de niveau de service (*Service Level Agreement* – SLA) agit comme une relation contractuelle et convenue qui comporte des aspects techniques, temporels, organisationnels et contextuels pour garantir la fourniture du service. Dans notre proposition, nous essayons d'aborder les trois premiers défis cités ci-dessus. Pour cela, dans la section suivante, nous identifions les principales contributions de notre travail par rapport à l'état de l'art. Ensuite, nous présentons la problématique et nous détaillons les trois parties principales de notre proposition, à savoir : la notion de service lui-même dans le contexte de SLA, la façon

7. http://liuppa.univ-pau.fr

dont nous associons notre proposition à la gestion des responsabilités et enfin, le modèle de réputation considéré.

5.2. État de l'art

La nécessité d'étudier les aspects sécurité dans les environnements inter-organisationnels est admise dans notre communauté de recherche, et des travaux tels que FI-OrBAC (Layouni, Pollet, 2009) et MultiOrBAC (El Kalam, Deswarte, 2006) en attestent. D'autres initiatives de sécurité inter-organisationnelle comme TOrBAC (El Kalam, Marzouk, 2012) et Multi-Trust-OrBAC (Ben Saidi, Marzouk, 2013) incluent dans leur politique de sécurité, comme dans la présente contribution, un modèle de réputation. Cependant, ces travaux ne prennent pas en compte la notion de contrôle d'usage des données, ni n'envisage de façon explicite le processus de négociation entre les différents partenaires impliqués dans le *workflow*. Un travail similaire à notre proposition est rapporté dans (Alnemr *et al.*, 2010). Les auteurs utilisent eux aussi les objectifs de niveau de service pour déterminer la réputation basée sur de critères liés à un contexte. Néanmoins, les principales différences sont qu'ils fondent leurs mesures des objectifs de niveau de service (*Service Level Objectives* – SLO) sur des conditions fonctionnelles traditionnelles. En outre, ils affirment que les informations de la réputation de chaque partenaire doivent être publiques et partagées entre les différents partenaires. Par ailleurs, dans la littérature, plusieurs propositions abordent le problème de l'utilisation des données par un troisième partenaire. Sans *et al.* (2007) proposent une approche de contrôle de l'usage en fonction de conditions contextuelles qui sont vérifiées pendant et après que l'autorisation ait été accordée. D'une manière générale, à notre connaissance, le contrôle d'usage a été défini et mis en œuvre comme un accès en continu (Pretschner *et al.*, 2009). En tenant compte de l'état actuel de l'évolution de nos domaines de recherche, voici les principales contributions de notre travail :

– compléter les techniques d'autorisation et d'authentification traditionnelles en passant de *qui* peut accéder à *quelles* ressources et *quand*, à *comment* les ressources sont utilisées une fois accessibles ;

– proposer une nouvelle approche de l'utilisation des SLO dans laquelle nous démontrons le phénomène de niveau de service propagé ;

– mettre en œuvre notre notion de niveau de service propagé pour créer des *workflows* dynamiques et pour déterminer un degré de réputation pour chaque partenaire.

5.3. Problématique et actions réalisées

Les SLA sont des documents signés, consentis et convenus qui engagent à la fois le client et le fournisseur du service sur le respect d'une liste de règles conditionnant la fourniture du service. Cependant, un des problèmes le plus importants est la difficulté de représenter ces conditions là liées aux aspects juridiques et à la logique métier dans un système d'information. Plusieurs approches ont été proposées pour combler l'écart

entre ces aspects « abstraits » et techniques. Il y a des langages de modélisation qui mettent en œuvre des contrats lisibles par une machine, comme SLANG (Lamanna *et al.*, 2003), WSLA (Keller, Ludwig, 2003) et WS-Agreement (Oldham *et al.*, 2006). Néanmoins, le problème persiste de savoir comment définir ce qu'est un niveau de service dans un environnement inter-organisationnel et les aspects qui doivent être représentés. Dans notre proposition, nous considérons que la notion de niveau de service inter-organisationnel est inexorablement liée à l'utilisation des données faite par le fournisseur de service. Ainsi, notre approche repose sur une nouvelle analyse du niveau de service pour maîtriser les données dans une environnement inter-organisationnel, tout en maintenant notre approche dans un cadre de gestion des responsabilités.

5.3.1. Notion de niveau de service

Le problème consiste à définir, dans un environnement inter-organisationnel, ce que signifie un accord de niveau de service. Ce n'est pas une question triviale, comme il est indiqué dans (Munier *et al.*, 2013), car dans ce genre d'environnements, bien que les partenaires aient convenu d'un objectif final commun, chacun d'eux possède néanmoins ses propres objectifs. Suivant notre approche orientée contrôle d'usage, nous définissons le niveau de service inter-organisationnel en termes de respect des conditions d'utilisation imposées par le propriétaire des données et représentées sous la forme d'objectifs de niveau de service (SLO).

Nous proposons également la notion de *niveau de service propagé* (*Propagated Service Level* – PSL) en tant que stratégie nous permettant de lier et de surveiller les SLO entre les différents partenaires impliqués dans le *workflow*, et plus important encore, d'adapter et de négocier les SLO en cas de non-conformité.

Pour illustrer le concept de PSL, prenons le cas d'une petite entreprise appelée ARMOR qui vend des ordinateurs assemblés sur mesure. ARMOR propose ses services sur Internet. C'est *via* son site Internet que les utilisateurs configurent leurs ordinateurs et paient pour le produit final. En interne, ARMOR met en œuvre deux politiques lui permettant d'économiser de l'argent. Premièrement, ARMOR n'a pas les produits en stock, ce qui signifie que l'entreprise sous-traite à une autre société les pièces d'ordinateurs en fonction de la demande. Deuxièmement, l'entreprise extérieure (FastSend INC) chargée du processus de livraison n'enlève les colis qu'une seule fois par semaine. Cet exemple représente un environnement inter-organisationnel qui implique l'acheteur du produit, ARMOR agissant en tant que vendeur, le fournisseur de pièces et FastSend INC pour la livraison (dans cet exemple, nous ne considérons pas la banque qui interagit également avec ARMOR). Supposons que dans le SLA initial établi entre l'acheteur et ARMOR, le délai de livraison convenu est de neuf jours. Pour une raison de logistique interne, le fournisseur ne peut pas fournir les pièces requises dans le délai fixé entre eux et ARMOR. Dans ce cas, ARMOR pourrait envisager les options suivantes :

– ARMOR accorde au fournisseur des pièces un délai supplémentaire et négocie avec l'acheteur une nouvelle date de livraison du produit final ;

– ARMOR attend la réception des pièces et négocie avec FastSend INC un enlèvement exceptionnel afin de respecter le SLA avec l'acheteur ;

– ARMOR attend la réception des pièces et recherche un autre transporteur offrant une livraison plus rapide ;

– ARMOR cherche un autre fournisseur pouvant lui livrer les pièces plus rapidement afin d'éviter de renégocier le délai de livraison avec FastSend INC.

Avec cet exemple simple nous illustrons le principe du niveau de service propagé, en faisant valoir que la non-réalisation d'au moins un SLO par un nœud (partenaire) du *workflow* génère un effet domino sur la totalité des SLO en déclenchant un processus de négociation. Avec notre approche de niveau de service propagé, nous considérons qu'avant que les données puissent être utilisées par un autre partenaire du *workflow*, celui-ci doit être informé de l'état du système et négocier les SLO non respectés. Cela implique que l'algorithme qui régit notre approche de PSL a en entrée les objectifs de service et produit en sortie un résultat final (si nécessaire) et la preuve de l'usage qu'il prétend avoir fait des données tout au long du *workflow*; le processus lui-même est considéré, par conséquent, comme dynamique et dépend complètement de la réalisation (ou non) des SLO.

Dans le schéma général de l'approche proposée, outre le processus principal, trois autres composants sont impliqués, à savoir, le coordinateur du *workflow*, le composant de négociation et le composant nommé *Universal, Description, Discovery and Integration (UDDI)*. La principale fonction du coordinateur est de mettre à jour l'état du système, de distribuer les fonctions aux nœuds, d'évaluer l'accomplissement des SLO et, en cas d'un comportement incorrect, d'informer le composant de négociation de la situation afin qu'il essaye, dynamiquement, de trouver dans l'UDDI un autre service qui pourrait garantir l'accomplissement des SLO au niveau global.

5.3.2. Niveau de service et gestion de la responsabilité

Comme il a été indiqué, la sortie du processus est un journal qui sert de preuve de l'usage qui a été fait des données à l'intérieur du *workflow*. Afin d'utiliser le journal comme preuve il faut qu'il ait les propriétés suivantes :

– il ne peut être modifié ou accessible que par des entités autorisées ;

– il doit détailler le flux des données complet, incluant la date à laquelle les données ont été utilisées, la raison de l'utilisation, l'action effectuée avec les données ainsi que le responsable de l'action ;

– il doit garantir que les informations enregistrées dans le journal sont fiables en signant l'entrée du journal.

Quant à la mise en œuvre, l'architecture pourra être centralisée ou distribuée, mais il faudra que la création du journal soit synchronisée avec l'exécution du *workflow*, et donc également avec le flux de données. Supposons que les données transmises soient modifiées par deux nœuds indépendants suite à un embranchement du *workflow*. Puisque le coordinateur du *workflow* connaît la configuration globale du système, il pourrait éviter les mises à jours incorrectes des données, ou les erreurs d'enregistrement ou de fusion dans le journal.

5.3.3. Niveau de service et réputation

Un point à prendre en compte est la façon dont le composant de négociation choisit un nouveau fournisseur de services pour l'intégrer au *workflow*. Nous utilisons pour cela la notion de réputation. La première fois qu'un client accède aux services d'un fournisseur, le client lui fait confiance quant à l'accomplissement des SLO convenus. Cependant, en cas de violation, le système applique une pénalité au fournisseur. Au niveau de l'architecture, la base de données contenant les informations sur la réputation de chaque entité ne pourra être accédée que par le composant de négociation afin de préserver l'intimité de l'entité, mais aussi parce que nous proposons une mesure subjective de la réputation basée sur des seuils et les sanctions établies par le client du service en tenant compte seulement de ses objectifs. Dans notre proposition nous attribuons des sanctions pour la violation de SLO liés à l'utilisation des données, mais aussi de la politique de sécurité.

5.4. Actions futures

Contrôler les données qui sont partagées ou fournies à un prestataire externe est un domaine de recherche actif dans le monde entier en raison des implications juridiques et commerciales. Dans cet article, nous avons introduit notre proposition de contrôle d'usage des données sur la base de relations contractuelles entre les entités manipulant ces données. Nous visons à établir une chaîne de responsabilités dans l'exécution d'un *workflow* en créant des journaux, mais aussi avec l'application d'une stratégie de sanction sur la base de la violation de la politique de sécurité et des objectifs de niveau de service convenus quant à l'utilisation des données. Nous proposons aussi l'applicabilité des niveaux de service propagés dans un *workflow* inter-organisationnel.

La prochaine étape consiste à intégrer et à formaliser les SLO en tant qu'élément de la politique de sécurité inter-organisationnelle. Notre idée initiale est orientée vers l'utilisation d'OrBAC ou d'une de ses variantes telle que Multi-Trust-OrBAC (Ben Saidi, Marzouk, 2013) et inclut de nouveaux éléments liés au but de l'utilisation des données, en proposant également une nouvelle catégorie pour les conditions contextuelles associées à l'utilisation des données dans un environnement inter-organisationnel.

Bibliographie

Agrawal R., Imieliński T., Swami A. (1993). Mining association rules between sets of items in large databases. In *SIGMOD'93: Proceedings of the International Conference on Management of Data*, p. 207–216. New York, NY, USA, ACM.

Agrawal R., Srikant R. (1994). Fast algorithms for mining association rules in large databases. In *VLDB'94: Proceedings of the 20th International Conference on Very Large Data Bases*, p. 487–499. San Francisco, CA, USA, Morgan Kaufmann Publishers Inc.

Alnemr R., Koenig S., Eymann T., Meinel C. (2010, 08). Enabling Usage Control through Reputation Objects: A Discussion on e-Commerce and the Internet of Services Environments. *Journal of theoretical and applied electronic commerce research*, vol. 5, p. 59–76.

Andonoff E., Hanachi C., Nurcan S. (2012). Adaptation des processus d'entreprise. In P. Lopisteguy, D. Rieu, P. Roose (Eds.), *L'adaptation dans tous ses états*, p. 78–122. Cépaduès Editions.

Ben Saidi M., Marzouk A. (2013). Multi-Trust-OrBAC: Access control model for multi-organizational critical systems migrated to the cloud. *International Journal of Soft Computing and Engineering*, vol. 3, n° 2, p. 51–55.

Brandenburg H., Wojtyna J. P. (2006). *L'approche processus : mode d'emploi* (2e éd.). Paris, Éditions d'organisation.

Bricon-Souf N., Dufresne E., Watbled L. (2005). Integration of the cognitive knowledge of activity in a service oriented architecture in the home care context. In *Studies in health technology and informatics*, vol. 116, p. 923–928.

Brézillon P. (2003). Using context for supporting users efficiently. In *HICSS'03: Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, p. 127.3. Washington, DC, USA, IEEE Computer Society.

Bubeck S., Cesa-Bianchi N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, vol. 5, p. 1-122.

Cabanac G. (Ed.). (2014). *Actes du 7e Forum Jeunes Chercheurs du congrès INFORSID.* Consulté sur http://www.irit.fr/~Guillaume.Cabanac/docs/fjc2014/fjc2014actes.pdf

Candillier L., Chevalier M., Dudognon D., Mothe J. (2012). Multiple similarities for diversity in recommender systems. *International Journal On Advances in Intelligent Systems*, vol. 5, n° 3, p. 234–246.

Chaari T., Zouari M., Laforest F. (2009). Ontology based context-aware adaptation approach. In D. Stojanovic (Ed.), *Context-aware mobile and ubiquitous computing for enhanced usability*, p. 26–58. IGI Global.

Chalmers D., Dulay N., Sloman M. (2004). Towards reasoning about context in the presence of uncertainty. In *Proceedings of the 1st international workshop on advanced context modelling, reasoning and management*, p. 1–33.

Collectif DGOS. (2012). *Recommandations pour la mise en œuvre d'un projet de télé-médecine. déploiement technique : Urbanisation et infrastructure.* Consulté sur http://esante.gouv.fr/sites/default/files/Recommandations_mise_en_oeuvre_projet_telemedecine.pdf

Collectif Inforsid. (2012). La recherche en systèmes d'information et ses nouvelles frontières. *Ingénierie des Systèmes d'Information*, vol. 17, n° 3, p. 9–68.

Conan D., Rouvoy R., Seinturier L. (2007). Scalable processing of context information with COSMOS. In J. Indulska, K. Raymond (Eds.), *LNCS: Distributed applications and interoperable systems*, vol. 4531, p. 210–224.

Davis R. (2001). *Business process modelling with aris: A practical guide*. London, Springer-Verlag.

Desrues L. (2008). *La télémédecine, un outil au service de l'amélioration du système de santé. l'exemple du réseau telurge dans le Nord - Pas-de-Calais*. Mémoire de Master non publié, École des hautes études en santé publique.

Dey A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing*, vol. 5, n° 1, p. 4–7.

El Kalam A. A., Deswarte Y. (2006, November). Multi-OrBAC: A New Access Control Model for Distributed, Heterogeneous and Collaborative Systems. In *8th IEEE International Symposium On Systems And Information Security*. São Paulo, Brésil.

El Kalam A. A., Marzouk A. (2012). TOrBAC: A trust organization based access control model for cloud computing systems. *International Journal of Soft Computing and Engineering*, vol. 2, nº 4, p. 122–130.

Gensel J., Villanova-Oliver M., Kirsch-Pinheiro M. (2008). Modèles de contexte pour l'adaptation à l'utilisateur dans des systèmes d'information web collaboratifs. In *Actes de l'atelier sur la Modélisation Utilisateur et Personalisation d'Interfaces Web à EGC'08: 8es journées francophones d'Extraction et Gestion des Connaissances*, p. 5–15.

Keller A., Ludwig H. (2003). The WSLA framework: Specifying and monitoring service level agreements for web services. *Journal of Network and Systems Management*, vol. 11, nº 1, p. 57–81.

Kohli P., Salek M., Stoddard G. (2013). A fast bandit algorithm for recommendation to users with heterogenous tastes. In *AAAI'13: Proceedings of the Twenty-Seventh Conference on Artificial Intelligence*, p. 1135–1141.

Koren Y. (2010). Collaborative filtering with temporal dynamics. *Communications of the ACM*, vol. 53, nº 4, p. 89–97.

Lacombe A., Boudy J., Baldinger J. L., Delavault F., Muller M., Farin I. *et al.* (2006). Telemedecine for elderly patient at home : the TelePat project. In *Proceedings of the international conference on smart homes and health telematics*, vol. 19, p. 74–81.

Lamanna D. D., Skene J., Emmerich W. (2003). SLAng: A language for defining service level agreements. In *FTDCS'03: Proceedings of the The Ninth IEEE Workshop on Future Trends of Distributed Computing Systems*, p. 100. Washington, DC, USA, IEEE Computer Society.

Layouni F., Pollet Y. (2009, February). FI-ORBAC : A Model of Access control for federated identity platform. In *IADIS 2009, the International Conference on Information System.* Barcelona, SPAIN. (ISBN: 978-972-8924-79-9)

Li L., Chu W., Langford J., Schapire R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *WWW'10: Proceedings of the 19th International World Wide Web Conference*, p. 661–670.

Montaner M. (2003). A taxonomy of recommender agents on the internet. *Artificial Intelligence Review*, vol. 19, nº 4, p. 285–330.

Morley C., Hugues J., Leblanc B., Hernandez J. A., Joly R., Gardian G. (2006). *UML 2 pour l'analyse d'un système d'information : Le cahier des charges du maître d'ouvrage* (3e éd.). Paris, Dunod.

Munier M., Lalanne V., Ardoy P.-Y., Ricarde M. (2013). Legal issues about metadata data privacy vs information security. *DPM'2013*, vol. 8247, p. 162-177.

Najar S., Kirsch Pinheiro M., Souveyet C. (2012). Mécanisme de prédiction dans un système d'information pervasif et intentionnel. In *UBIMOB'12: Actes des 8es journées francophones mobilité et ubiquité*, p. 146–157. Cépaduès Editions.

Noyé D. (1997). *L'amélioration participative des processus* (2e éd.). INSEP.

Nurcan S. (2008). A survey on the flexibility requirements related to business processes and modeling artifacts. In *HICSS '08: Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, p. 378. Washington, DC, USA, IEEE Computer Society.

Oldham N., Verma K., Sheth A., Hakimpour F. (2006). Semantic WS-agreement partner selection. In *WWW'06: Proceedings of the 15th International Conference on World Wide Web*, p. 697–706. New York, NY, USA, ACM.

OMG. (2010). *Business process model and notation (bpmn 2.0)*. Rapport technique n° 2. Object Management Group.

Perréal F. (2003). Gluconet: Télémédecine et suivi thérapeutique du patient diabétique. In *Actes des 12es journées des technologies de communication hospitalière.* Montpellier.

Pignotti E., Edwards P., Grimnes G. A. (2004). Context-aware personalised service delivery. In *ECAI'04: Proceedings of the European Conference on Artificial Intelligence*, p. 1077–1078.

Pretschner A., Schütz F., Schaefer C. (2009). Policy evolution in distributed usage control. *Electronic Notes in Theoretical Computer Science*, vol. 244, p. 109-123.

Priss U. (2006). Formal concept analysis in information science. *Annual Review of Information Science and Technology*, vol. 40, n° 1, p. 521–543.

Radlinski F., Kleinberg R., Joachims T. (2008). Learning diverse rankings with multi-armed bandits. In *ICML'08: Proceedings of the 25th International Conference on Machine Learning*, p. 784–791.

Ricci F., Rokach L., Shapira B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook*, p. 1–35. Springer.

Rubens N., Kaplan D., Sugiyama M. (2011). Active learning in recommender systems. In *Recommender systems handbook*, p. 735-767. Springer.

Sakka E., Prentza A., Lamprinos I. E., Leondaridis L., Koutsouris D. (2004). Integration of monitoring devices in the e-vital service. In *Proceedings of the 26th annual international conference of the ieee embs*, vol. 4, p. 97–100. San Francisco, CA, USA.

Sans T., Cuppens F., Cuppens-Boulahia N. (2007). A framework to enforce access control, usage control and obligations. *Annales Des Télécommunications*, vol. 62, n° 11-12, p. 1329-1352.

Schilit B. N., Theimer M. M. (1994). Disseminating active map information to mobile hosts. *IEEE Network*, vol. 8, n° 5, p. 22–32.

Schonenberg H., Mans R., Russell N., Mulyar N., Aalst W. Van der. (2008). Process flexibility: A survey of contemporary approaches. In *International workshop on ciao/eomas, held at international conference on advanced information systems*, p. 16–30. Montpellier.

Sigg S., Haseloff S., David K. (2010). An alignment approach for context prediction tasks in ubicomp environments. *IEEE Pervasive Computing*, vol. 9, n° 4, p. 90–97.

Sutton R. S., Barto A. G. (1998). *Introduction to reinforcement learning* (1re éd.). Cambridge, MA, USA, MIT Press.

Vanrompay Y., Kirsch-Pinheiro M., Berbers Y. (2011). Service selection with uncertain context information. In S. Reiff-Marganiec, M. Tilly (Eds.), *Handbook of Research on Service-Oriented Systems and Non-Functional Properties: Future Directions*, p. 192–215. IGI Global.

Weber B., Reichert M., Rinderle S. (2008). Change patterns and chage support features – Enhancing flexibility in process-aware information systems. *Internationnal Journal on Data and Knowledge Engineering*, vol. 66, nº 3, p. 438–466.

Weber B., Sadiq S., Reichert M. (2009). Beyond rigidity-dynamic process lifecycle support: A survey on dynamic changes in process-aware information systems. *International Journal on Computer Science – Research and Development*, vol. 23, nº 2, p. 47–65.

Wen Q., He J. (2006). Personalized recommendation services based on service-oriented architecture. In *APSCC'06: Proceedings of the IEEE Asia-Pacific Conference on Services Computing*, p. 356–361.

Weske M. (2012). *Business process management*. Berlin, Springer-Verlag.

Wille R. (2005). Formal concept analysis as mathematical theory of concepts and concept hierarchies. In B. Ganter, G. Stumme, R. Wille (Eds.), *LNCS: Formal concept analysis*, vol. 3626, p. 1–33. Springer.